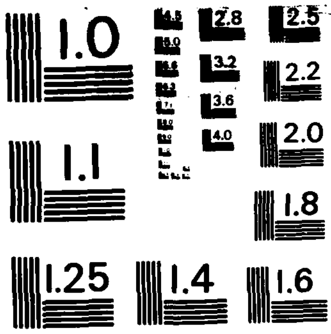
AD-A139 874 EVIDENCE ACCUMULATION FOR SPATIAL REASONING IN AERIAL IMAGE UNDERSTANDING. (U) MARYLAND UNIV COLLEGE PARK CENTER FOR AUTOMATION RESEARCH V S HWANG ET AL. OCT 83 1/1

UNCLASSIFIED CAR-TR-28 AFOSR-TR-84-0178 F49620-83-C-0082 F/G 20/6 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

(4)

TR-28
:-1336

F-49620-83-
Octob

Evidence Accumulation for Spatial Reasoning in Aerial Image Understanding

Vincent Shang-Shouq Hwang
Takashi Matsuyama*
Larry S. Davis
Azriel Rosenfeld

CENTER FOR AUTOMATION RESEARCH

UNIVERSITY OF MARYLAND
COLLEGE PARK, MARYLAND
20742

(4)

CAR-TR-28
CS-TR-1336

F-49620-83-C-0082
October 1983

Evidence Accumulation for Spatial Reasoning in Aerial Image Understanding

Vincent Shang-Shouq Hwang
Takashi Matsuyama*
Larry S. Davis
Azriel Rosenfeld

Center for Automation Research
University of Maryland
College Park, MD 20742

Abstract

This document discusses

~~We describe~~ a control structure for building an Image Understanding System. This system can deal with objects with diverse appearances when consistent spatial relations exist between objects. By accumulating consistent predictions originated from existing instances, our system can dynamically reason about what to do in order to construct interpretations of the image. In this paper, we ~~have~~ discussed parts of the proposed system - the representation of spatial knowledge, the accumulation of evidence, the focus of attention mechanism, and the integration of constraints for top-down control.

*Permanent address: Department of Electrical Engineering
Kyoto University, Sakyo, Kyoto, Japan

1. Introduction

1.1. Problems in Image Understanding

In image understanding, an image is given to a computer as input and the desired output is a labeled picture or a symbolic description of the image. In order to do this, the computer needs to have the knowledge about the scene to be described and needs to be able to use such knowledge to construct the description.

The following are some problems in the building of an image understanding system(IUS) that have not yet been treated successfully.

(1) Segmentation

An IUS needs to extract image features from the image. To do so, it needs to choose image processing methods to apply to the image. The method selected must be appropriate, e.g. cheap and effective. How to select appropriate image operators is a basic problem.

There are many methods of segmenting an image to extract objects. For example, thresholding, region growing, or specialized blob finding can be used to extract regions in an image. Each operator has its advantages and disadvantages. Using appropriate segmentation methods can increase the efficiency and reliability of the system.

(2) Diversity in Appearance

Most of the cultural structures in aerial photographs have many diverse appearances. An IUS needs to know which appearance description to search for in the image. One could let the IUS try every possible appearance description, but this is

not desirable, since the number of alternatives may be very large. How to limit the number of possible appearances and intelligently select the ones to try is another problem.

For example, houses in a suburban housing development have many possible shapes, sizes, and colors. We must know what type of house we are looking for when we are searching for houses. Elimination of search for unlikely appearances can increase the performance of the IUS.

(3) Representation and Manipulation of Domain Related Knowledge

An IUS needs to have domain related knowledge in order to construct an interpretation of the image. In our domain, the sources of knowledge are diverse and redundant. Requirements that must be satisfied by an object are specified in many ways, and each of them gives only a weak constraint. How to represent and manipulate domain knowledge is another problem.

For example, a house in a suburban residential area can be specified by its shape, size, and color as well as by its relations to other houses and roads. Each of these constraints specifies some requirements of a house. Knowing that only some of the constraints for a house are satisfied is not enough to assign the house label to a pictorial entity. On the other hand, failure to satisfy some of the constraints doesn't indicate that the pictorial entity can't be a house. Instead, it may indicate that further investigation is needed. A production rule based representation is not enough in our domain. A better representation method and control mechanism are needed in this domain.

1.2. Previous Work in Image Understanding

Much research has been done in the field of image understanding. In this section, we review a few of the existing image understanding systems.

Selfridge[Self82] developed a system to locate houses and roads in aerial photographs. He uses a technique called "reasoning about success and failure". His system uses information such as the shapes and sizes of regions and evaluates the performance of operations derived from explicit goals and explicit intensity data. Resegmentation is accomplished by changing the parameters of the image operators. Knowledge about how to adaptively change these parameters is represented by procedures. Spatial relations between objects are simple(e.g. adjacency).

Nagao and Matsuyama[Naga80] built a system that analyzes aerial photographs by assigning labels to regions. A color aerial photograph is first segmented into regions using several general image processing methods. Regions are characterized by their dominant features and specialized feature extraction and recognition programs are applied to appropriate regions. Knowledge about the assigning of labels to regions is represented by production rules. When several labels are assigned to a region, the system resegments the image by splitting or merging regions based only on the intrinsic properties(e.g. intensity, shape) of the regions.

Ohta[Ohta80] constructed a system to analyze outdoor scenes. It uses bottom up and top down analysis during the interpretation process. A color image is first segmented into regions. Many pieces of the image are identified and labeled during the bottom up processing using only intrinsic properties of regions. Semantic constraints between labeled regions are checked by a top down process. When major

changes are made to the already labeled regions during the top down analysis, bottom up analysis is reactivated to reevaluate the change. Domain knowledge is represented by production rules.

1.3. Important Issues In the Building of an IUS

Three issues are discussed here that are important in building an image understanding system.

(1) Knowledge Based Segmentation

It is advantageous to use a knowledge based segmentation system to process an image. Many studies have been done on picture processing operators. Their characteristics have been studied, such as effectiveness in extracting given types of pictorial entities in a given environment, required cost of processing, and possible artifacts caused by the operators. A knowledge based segmentation system uses such knowledge about the operators.

In such a system, a picture is a collection of pixels. The objects to be extracted are composed of sets of pixels. Picture processing operators are processes that group pixels into meaningful sets. Knowledge about the characteristics of image processing methods is used in the selection of methods. The aim of the system is to find methods which are cheap and are able to group pixels into desirable sets by reasoning about descriptions of the goal and environment and the characteristics of the operators.

For example, if we know the object has high contrast with the background, we would use thresholding rather than region growing, since this method is cheap and

effective in the given environment. On the other hand, if we know the picture is noisy and complicated, we must use a more sophisticated method to extract objects, since a simple thresholding method would not work well.

(2) Evidence Accumulation

An IUS builds interpretations and searches for missing objects in the image. Objects found(instances) can be used to predict missing objects(hypotheses). Hypotheses from various sources can be combined to guide the searching process. Such accumulation of evidence from different sources decreases the total amount of effort spent in processing and increases the reliability of the analysis.

In our domain, the spatial relations among objects are consistent. These relations are constrained by the functional purposes of the objects. For example, driveways function as linkages between roads and houses. This functional purpose constrains the spatial relations among these three objects. If a house is found, it can create hypotheses about the existence of roads and driveways around it. Many of these hypotheses, originating from different instances(house, road, or driveway), can be combined to indicate regions most likely to contain objects.

(3) Model Selection based on Contextual Information

When an IUS searches for a missing object in a region, it should use contextual information to predict the most likely appearance(s) of the object.

Let us assume that we have found a piece of road in a region. Suppose now we want to find a piece of road which is adjacent to the existing piece of road. We need to decide what is the exact appearance of the piece of road we are looking for before

we search for it. From our road knowledge, we know that road pieces which are adjacent to each other usually have the same width. This piece of knowledge and the contextual information lead us to look for a road piece which has the same width as the one already found.

1.4. A Control Structure for Image Understanding Systems

In this paper, we propose a control structure for building an image understanding system(see Figure 1.1), and apply it to the analysis of an aerial photograph of a suburban area containing houses, road, and driveways.

There are three levels of representation and analysis in the system: A High Level Reasoning Expert(HLRE) utilizes a symbolic hierarchical model for the possible spatial organizations of objects in the image to build partial, local interpretations of the image and to reason about where to further analyze the image and what analyses to perform. A Model Selection Expert(MSE) reasons on the basis of contextual information provided by the HLRE and selects the most promising appearance descriptions to use in searching for objects and structures in the image. A Low Level Vision Expert(LLVE) finds pictorial entities that satisfy these appearance descriptions by selecting effective image processing methods to find the appropriate entities.

Knowledge about objects is represented at several levels of specificity. For example, "house" is a generalization of many specifically shaped types of houses(e.g. rectangular or U-shaped). The HLRE determines the general class of objects to search for(e.g. house) while the MSE determines which specialization(e.g. rectangular) should be looked for. As illustrated in Figure 1.1, a common knowledge base is used by HLRE and LLVE to support their cooperation in deciding on the most

appropriate appearance.

We are currently concentrating on the design of the High Level Reasoning Expert, emphasizing the representation of domain knowledge and mechanisms for the accumulation of evidence and focus of attention. Both the Model Selection Expert and the Low Level Vision Expert are currently being simulated by a human.

2. High Level Reasoning Expert

2.1. Introduction

In this section, we discuss the principal technical issues in the design of the HLRE - the representation of knowledge, the representation of spatial relations, the accumulation of evidence, the focus of attention mechanism, and the intergration of constraints for top-down control of the MSE(situation selection).

2.2. Knowledge Representation for Objects

The appearances of objects in our domain are diverse. This diversity is currently handled by adopting a frame-based representation for object representation.

A frame is a data structure for a stereotyped object that is composed of "slots"[Fahl79,Mins75]. Information stored in the slots includes features of the objects and their relations to other objects. Default value assignments and attached procedures for slots are typical characteristics of a frame-based knowledge representation.

Frames are organized into a hierarchical structure by "part-of" relations. A frame at a higher level is an abstraction of lower level frames. For example, the "house unit" and "driveway" frames are members of the "house group" frame. They are linked to the "house group" frame by "part-of" links.

In addition to the "part-of" relation, every frame has two other slots for standard relations. The first one is the "a-kind-of" slot. When a frame is instantiated(i.e., when an instance of the entity represented by the frame is detected in an image), the

instance is represented by a frame and is linked to its prototype frame through the "a-kind-of" link. Properties of the frame are inherited by the instance through this link. Usually, there are many possible appearances for an object. Each appearance is a specialization of the general frame and is also linked to the frame by the "a-kind-of" link. When a frame is instantiated, one of the possible appearances is instantiated. However, knowledge about other possible appearances is accessible to the instance through its "a-kind-of" link. Figure 2-1 shows the "part-of" relation between the driveway, house, and house unit frames. Possible appearances for the shape of the house are linked to the house frame by an "a-kind-of" link. Instance H1 is instantiated as a rectangular house. It is linked to the rectangular house frame by an "a-kind-of" link.

The second standard slot is the "dependent" slot. During the interpretation process, existing instances are used to construct more complete partial interpretations. The newly derived interpretations are said to be dependent upon those existing instances which were used during the derivation process. If the features of some instances subsequently change, the features of other instances which depended on those instances should be checked, since such changes may affect the validity of the relations. In our system, the "dependent" link is used for this purpose and is used to chain the dependency of reasoning results.

A frame has many other slots; these slots can be used to store features of the object and methods for computing them.

2.3. Representation of Spatial Relations

In our system, binary spatial relations between specific classes of objects are described by computational procedures. Each procedure specifies an area relative to the first object in which the second object, referred to as the "target" object, must occur for the relation to hold. When a spatial relation is used to construct a prediction about the likely presence of other image structures, we call this area a "prediction area". In addition to this area specification, a set of constraints on the target object are also associated with a spatial relation. They describe the constraints that the target object must satisfy. Again, when the spatial relation is used to construct a prediction, these constraints are used by the MSE to choose a likely appearance model.

For example, in Figure 2-2, suppose R represents the area where one of the neighboring houses of house H0 should reside. Also, suppose H1 is a house. Since H1 overlaps with region R and this overlapping is significant, and H1 also satisfies the constraints for a house, house H1 is said to be a neighboring house of house H0. In our system, such a relation is recorded by storing house H1 in the "neighboring houses" slot of house H0.

2.4. Evidence Accumulation

Evidence concerning the existence of yet undiscovered structures can be obtained from hypotheses(predictions) that the Image Understanding System has constructed, but not yet verified, as well as from existing instances. When several prediction areas originating from different objects overlap, we accumulate the constraints from the contributing sources of evidence associated with the overlapping

regions and construct *contextual cues* for the MSE.

In our system, we currently only accumulate the constraints for the same type of object. The construction of the contextual cues is discussed in Section 2.6; here, we focus on the spatial data structure that supports the recognition of potentially supporting sources of evidence.

As an example, consider Figure 2-3,and suppose we have found road pieces R1 and R2. Each road piece can serve as a source of evidence concerning the presence of adjacent road pieces. Let E1, E2, E3, and E4 be the associated predictions. As we can see, E2 and E3 overlap. Let the overlap region be O. Then we can say that our confidence in finding a road piece near region O increases, since it is supported by both E2 and E3.

Let us examine another example. In Figure 2-4, suppose two road pieces R1 and R2 have been found in the image. As usual, each road piece creates predictions E1, E2, E3, and E4 concerning potential adjacent road pieces. The prediction area of E3 overlaps with that of E1. However, the constraint on the direction of the road imposed by E1 differs from the constraint imposed by E3 is so that our confidence about finding an intersection near region O increases.

These examples suggest that both instances and hypotheses should be represented using a *symbolic/iconic* data structure that associates highly structured symbolic descriptions of the instances and hypotheses with regions in an array. The regions are represented by bit planes having 1's at pixels of the region and 0's elsewhere.

2.5. Focus of Attention

All sources of evidence, instances and hypotheses, are recorded in a common database, as discussed in Section 2.4. Our focus of attention mechanism is a sequential control structure that prioritizes consistent sets of sources of evidence(instances and predictions) and pursues the most likely consistent set.

Consider as an illustrative example Figure 2-5. E1 overlaps with E2. This overlap suggests the existence of an intersection at O1. However, the overlap of E1 and E3 suggests the existence of a connecting road piece between R1 and R3.

Define a *situation* as the collection of all mutually consistent evidence along with a region(called the *validity region*) in which the situation might obtain. A situation can arise from interactions between instances and hypotheses. The system can focus its attention only on situations.

During the interpretation process, the system needs to select a situation with a good expectation. It does so on the basis of a measurement of its belief in the situation.

Let situation S be due to the accumulation of evidence from E1, E2, ... En and let Di be the confidence measure for evidence Ei. Then we define the confidence measure of S as the summation of the Di's.

In general, one can imagine selecting more than one situation for analysis, and processing them independently and simultaneously. However, we do not have good criteria to determine if two situations are independent of each other, since as a result of analyzing a situation, the common database may change. New instances may be inserted into the database, and attributes of other sources of evidence may change. If

two situations S1 and S2 are selected, when the system resolves situation S1, situation S2 may no longer exist. Some hypotheses that participated in S2 may have been canceled or disproved while the system resolved S1. Identifying situations which can be processed independently is a topic for future research. In our implementation, we process only the situation with best confidence measure among all situations.

2.6. Resolving Situations

This section discusses the computational mechanism used by the system for developing *contextual cues* for the MSE. Contextual cues are constructed independently by each instance of the situation chosen by the focus of attention mechanism.

In Figure 2-6, suppose hypothesis E1 is created by road piece R1. Hypothesis E1 overlaps with road piece R2. In this overlapping region the constraints from E1 and R2 are accumulated and a situation is created. In order to resolve this situation, the system asks R1 if the situation's context satisfies the expectations that led it to predict E1. Here, R1 would not be satisfied, since R2 is not *adjacent* to it. R1 would then direct the MSE to find a road piece which joins R1 and R2. The directions are contained in a *basic action* which is a directive to the MSE constructed by the instance. Each basic action is a 4-tuple(Goal, Region, Contextual Cue, and Level of Effort). The *Goal* attribute indicates what instance to search for while the *Region* attribute indicates where to search. The *Contextual Cue* attribute contains the context information computed by the instance to be used by MSE in deciding what appearance(s) to search for. The level of effort the instance allows the MSE to use to execute this action is recorded in the *Level of Effort* attribute. When an instance creates a basic action, it uses the context information recorded in the currently

focused situation and the attached knowledge of the instance to construct the basic action.

Figure 2-7 shows a basic action. It represents a request to MSE to find a road instance from point A to B inside region O with high effort. The width and the orientation of the road instance to be found are 0 degree and 10 pixels respectively.

For each selected situation, many basic actions are usually generated. Each of them is constructed independently by the instances contributing to the establishment of a situation. HLRE must establish an order for executing the basic actions. Also, some basic actions may be redundant, since similar basic actions can be constructed by different instances examining the same situation. HLRE should summarize similar basic actions so that MSE examines only those basic actions which are necessary. We are currently studying this topic.

2.7. Interpretation Process

Initially, a given set of image processing operators is applied to the image to construct a set of segments that are interpreted by the HLRE to form the initial set of instances. The system then iteratively performs a process of hypothesis formation → situation construction → situation resolution(through the focus of attention mechanism) → hypothesis formation ... until a stage is reached where all hypotheses have been pursued to their ultimate conclusions. The system currently does not prune hypotheses and situations as they become unlikely(which it should), but it does dynamically reorder situations and edit actions based on new instances constructed during the interpretation process.

3. Experimental Results

The image used in our experiment is a 320 by 160 portion of an aerial image(Figure 3-1). The intensity at each pixel ranges from 0 to 63. The scene contains houses, roads, trees, and driveways.

The appearance models we are using are a subset of the possible models for suburban housing developments. Currently, we deal only with the houses, road pieces, road intersections, and the spatial relations among them. A house may have many possible prototypes(e.g. rectangular, U-shaped). In the current implementation, we only use the rectangular prototype. Figure 3-2 shows the default constraints for a house and the spatial relations between a house and other related objects. The prototype for a road piece is described by an elongated rectangle. It has spatial relations to other adjacent road pieces and adjacent road intersections. Figure 3-3 shows the knowledge about a road piece and its spatial relations to other objects. A road intersection is modeled by a rectangle. It is the intersection of two road pieces which intersect at a sufficiently sharp angle.

The system's analysis starts with the segmentation of the image. Since the houses and road pieces are modeled by compact and elongated rectangles, such rectangles are first extracted from the image. A simple blob finder and ribbon finder are used to find blobs and ribbons in the image.

Compact rectangles are initially instantiated as house instances and elongated rectangles as road piece instances. These instances constitute the initial entries in the iconic database. Figures 3-4 and 3-5 show the initial house instances and road piece instances extracted from the image. As we can see, some areas of the image are inter-

preted as both house and road.

Now, the interpretation process starts. In the first cycle, the system checks each instance and, for each spatial relation, creates a hypothesis, if possible, and inserts it into the database. Since some of the spatial relations may depend on yet undetermined values stored in frame slots, not all spatial relations may be hypothesized at the beginning(unless the default values for these slots are sufficiently reliable).

Figure 3-6 shows all the instances and hypotheses of houses in the database. House instances are indicated by white solid rectangles while house hypotheses are indicated by hollow rectangles. Figure 3-7 shows all the instances and hypotheses of road pieces.

In the second cycle, the system's focus of attention mechanism selects the situation with best context information. Currently, we use the number of pieces of supporting evidence as a measure to compute the merit of a situation.

Figure 3-8 shows a situation selected by the system. It has four pieces of evidence supporting the existence of a road. The white solid region indicates the overlap region of these four sources of evidence. The hollow rectangles indicate the instances and hypotheses participating in the situation.

The instances participating in this situation are road pieces R1, R2, and houses H1, H2. A situation is represented by a frame with two slots - direct evidence and indirect evidence. The indirect evidence slot contains all those instances whose hypotheses contributed to the formation of the situation while the direct evidence slot contains the instances which contributed directly. The situation in the current example is represented as follows:

indirect evidence : H1, H2, R1
direct evidence : R2

The system asks each instance participating in the situation to review what is currently known about the situation and to decide whether its prediction is validated or invalidated by the current knowledge. Here, H1 and H2 are satisfied with the current situation, since there is a road piece instance partially overlapping the validity regions of their hypotheses. In this case, no further action is required. In the case of road piece R1, however, the constraints are only partially satisfied. Road piece instance R2 fails to satisfy the adjacency constraint demanded by R1. A basic action is constructed by R1 to find a connecting road piece in region O. Road piece R2 has no hypothesis that needs to be validated(since it is a "direct instance"), therefore no constraints need to be satisfied, and R2 does not construct any basic action.

Since the system currently does not support any summarization process for determining redundancy among basic actions, the MSE must check to see if the execution of previous basic actions has produced new instances which would make the execution of its currently chosen action unnecessary.

In the current experiment, the MSE is simulated by a human. The descriptions of the action and the situation are displayed on the screen. The description of the result is entered from the terminal. The result obtained from the terminal is instantiated as an object instance and returned to the system.

Figure 3-9 shows another situation selected by the system. Figure 3-10 shows all the house instances in the database when the situation is selected. This situation has two pieces of evidence supporting it:

indirect evidence : H1, H2
direct evidence : none

No instance participates in this situation directly. The hypotheses that originated from H1 and H2 overlap at region O.

Now the system resolves the situation. Since no instance has been found in the area of interest, houses H1 and H2 request further analysis(since both H1 and H2 demand the existence of a house). A region where further analysis is to be done (region O in both cases) is computed by both H1 and H2 using the knowledge about houses and the context information associated with the situation.

Suppose the system first executes the basic action generated by H1. Since there is no house instance in region O, the system gives the selected action and the situation to the MSE. These descriptions are displayed on the screen. Finally, the result,i.e. the description of the object found in region O, is obtained from the terminal. The MSE instantiates it as a house instance and returns it to the system.

Since the other basic action has not yet been interpreted, the system checks to see if it needs to be executed. The house instance in region O detected as a result of executing the basic action from H1 makes it unnecessary to execute the basic action selected.

4. Conclusion

4.1. Discussion

In this paper, we have described a control structure for the building of an Image Understanding System. This system differs from many existing systems in its ability to represent and manipulate knowledge about objects with diverse appearances when consistent spatial relations exist between objects. It dynamically selects most likely appearances to search for, and adaptively chooses appropriate segmentation methods to process the image.

A frame-based method is used to represent domain related knowledge. Many type of links(e.g. part-of, a-kind-of, and spatial relations) exist between frames. To manipulate such knowledge, our system is decomposed into three different modules(HLRE, MSE, and LLVE) each of which uses different portions of the knowledge to do its task. Contextual cues collected by one module are used by another module to perform more efficient and more effective reasoning.

Our system constructs all consistent interpretations during the process. This can be very inefficient when the number of consistent interpretations is very large. However, by using enough knowledge about the domain objects, we believe the number of consistent interpretations can be kept small.

4.2. Work to be Done in the Future

We have currently implemented only parts of the proposed system - the representation of spatial knowledge, the accumulation of evidence, the focus of attention mechanism, and the intergration of constraints for top-down control of the

MSE. The following are some of the important issues that need to be studied.

Objects are organized into a hierarchical structure by "part-of" links. Sets of parts that satisfy particular spatial relations can be grouped together and can then be referred to as a unit. Although this grouping ability allows more efficient knowledge representation, it is not clear how this affects the evidence accumulation mechanism. For example, parts can have spatial relations with objects belonging to the same structural hierarchy as well as with objects belonging to a different structural hierarchy. When we group parts together, the resulting group can also have spatial relations to the same objects that the component objects had relations with. Should the whole and the parts it contains be treated as different sources in the accumulation of evidence?

One characteristic of our system is that it makes use of the least commitment principle. It constructs interpretations whenever no counterarguments are presented. An IUS can construct interpretations that are "ambiguous". For example, when our system establishes a link between two instances, there may not be enough contextual cues available to make the decision. As a result, the system can make incorrect decisions. How to recover from an incorrect decision made by the system is a topic to be studied in the future.

In our system, instances construct lists of basic actions to be executed by MSE. However, some of these basic actions can be redundant, since different instances participating in a situation may require similar basic actions to be performed. Also, some of the basic actions may be executed independently. From the efficiency point of view, our system should summarize the basic actions to identify redundant and

independent ones. How to do this deserves further study.

REFERENCES:

[Fahl79]

Scott E. Fahlman, "NETL, A System for Representing and Using Real-World Knowledge", MIT Press, Cambridge, Mass., 1979.

[Mins75]

Marvin Minsky, "A Framework for Representing Knowledge", in The Psychology of Computer Vision, Edited by P. H. Winston, McGraw Hill, New York, 1975, pp. 211-277.

[Naga80]

M. Nagao and T. Matsuyama, "A Structural Analysis of Complex Aerial Photographs", Plenum Press, New York, 1980.

[Ohta80]

Y. Ohta, "A Region-Oriented Image-Analysis System by Computer", Ph. D thesis, Kyoto University, 1980.

[Self82]

P. G. Selfridge, "Reasoning about Success and Failure in Aerial Image Understanding", Ph.D thesis, University of Rochester, 1982.

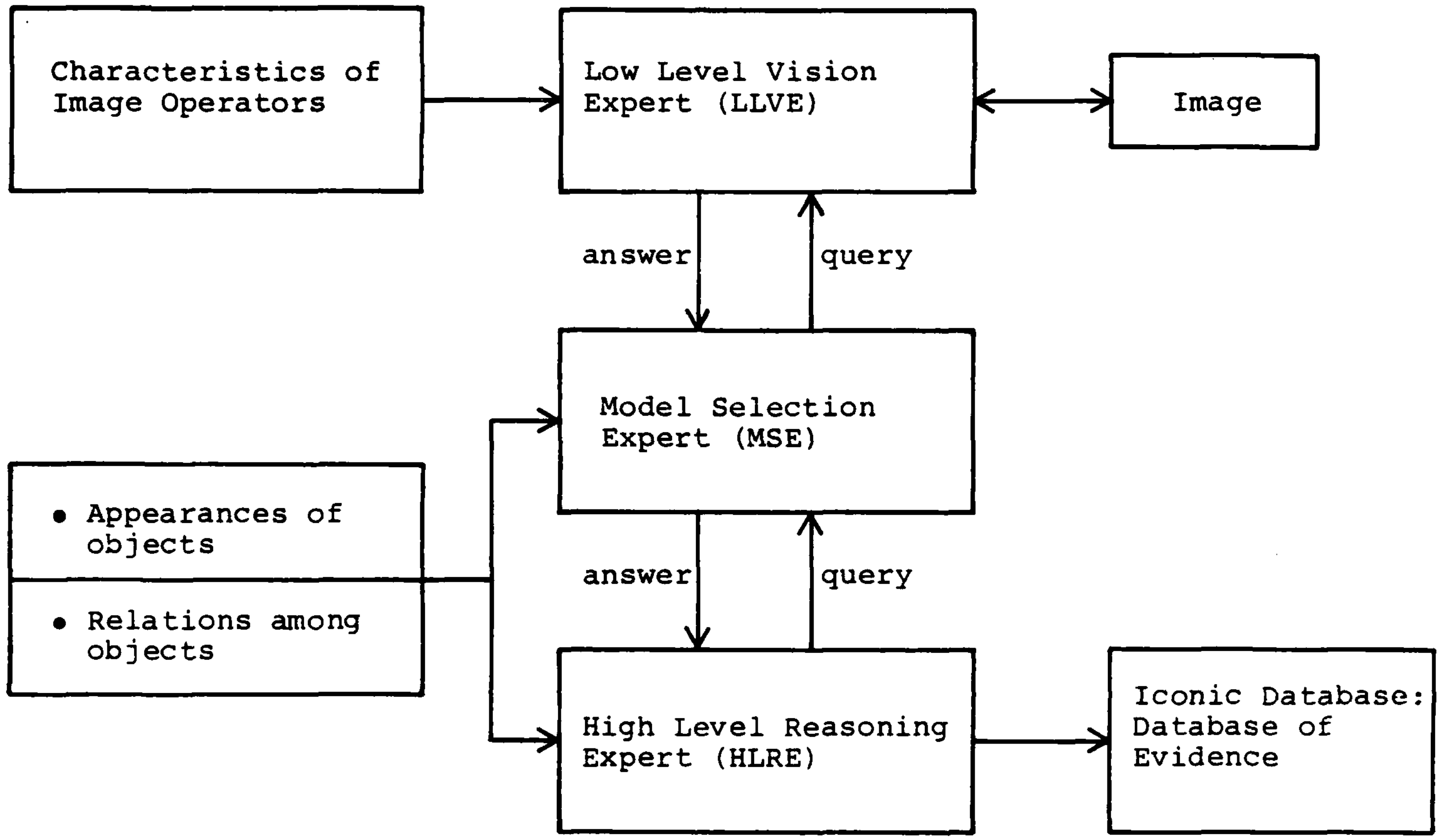


Figure 1-1. An Image Understanding System

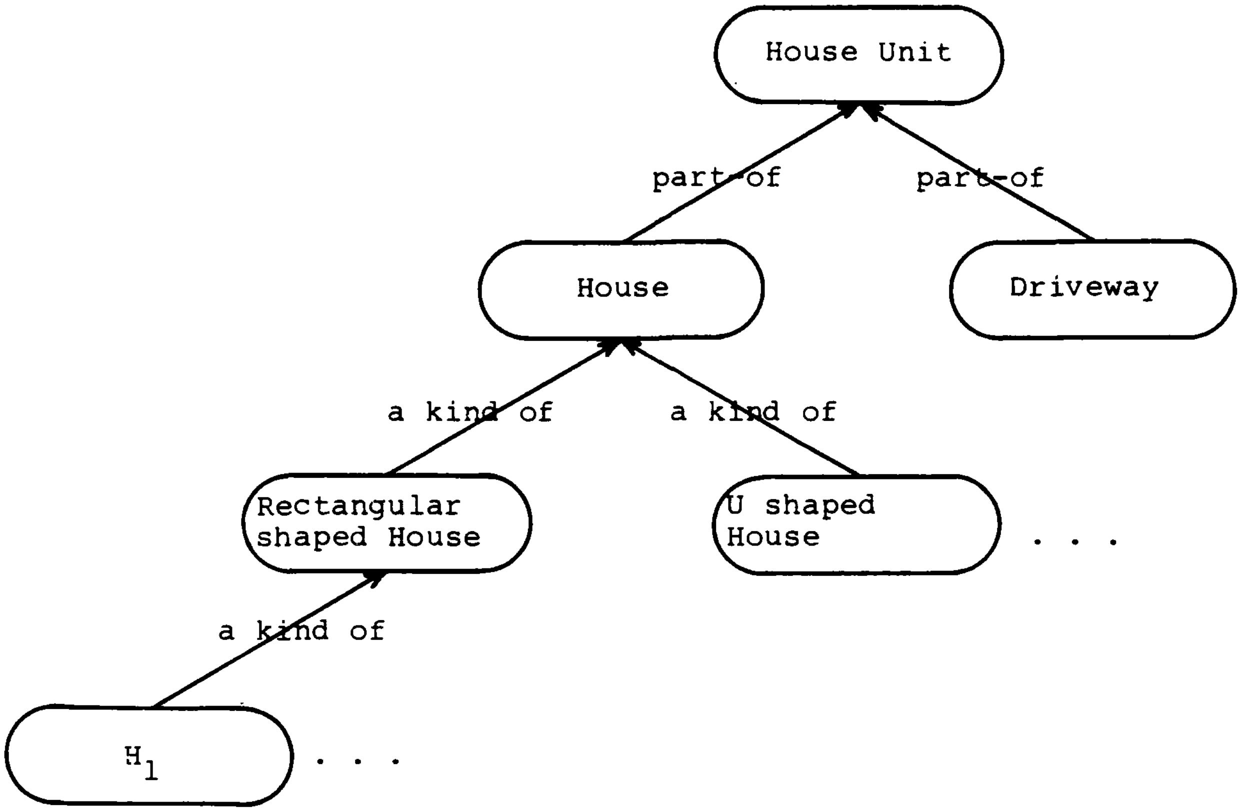


Figure 2-1. Part-of and a-kind-of relations.

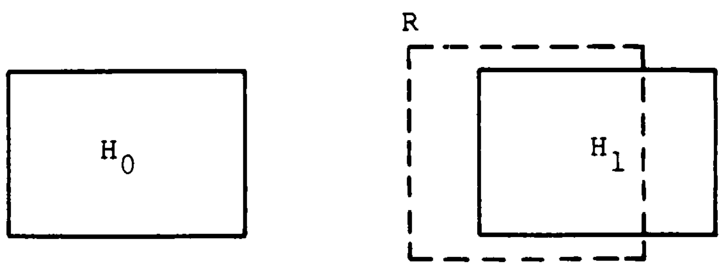


Figure 2-2. Adjacent neighboring house relation.

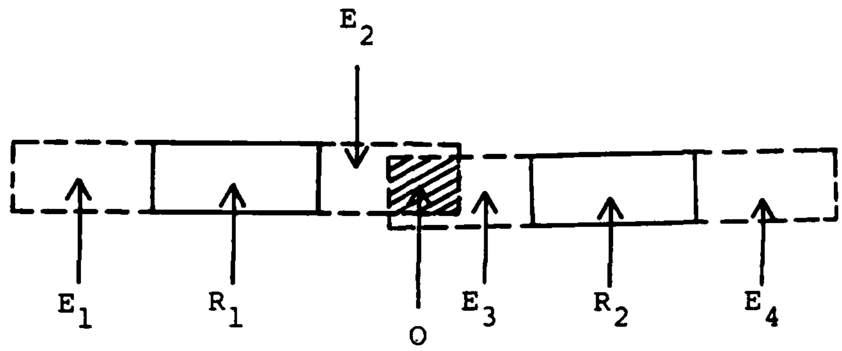


Figure 2-3. Two pieces of evidence overlap - prediction of connecting road.

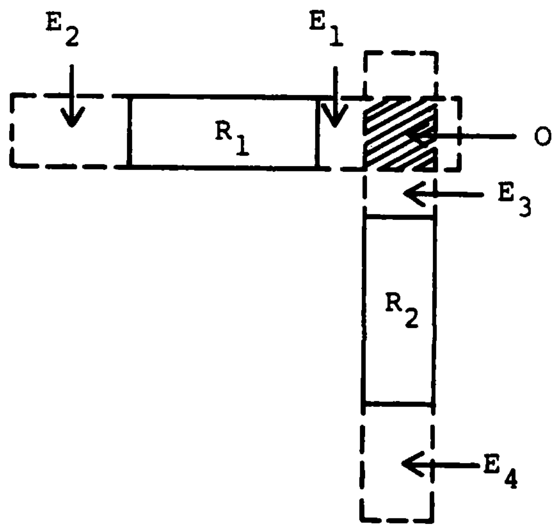


Figure 2-4. Two pices of evidence overlap - prediction of a road intersection.

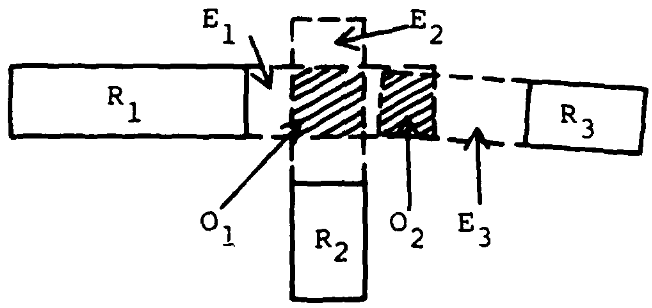


Figure 2-5. A piece of evidence can overlap with many other pieces of evidence.

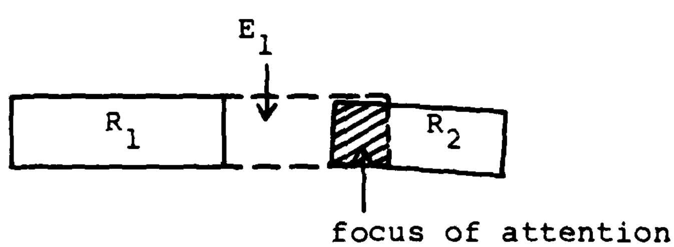


Figure 2-6. Resolving a situation.

Goal: find road piece

Region: region 0

Contextual cues: (a) from point A to point B
(b) width = 10 pixels
(c) orientation = 0 degrees

Level of effort: high

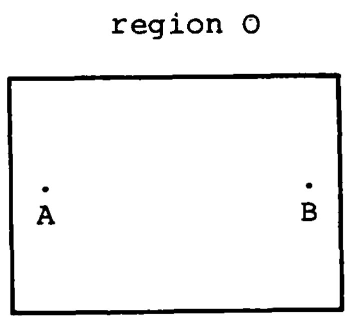


Figure 2-7. A basic action.

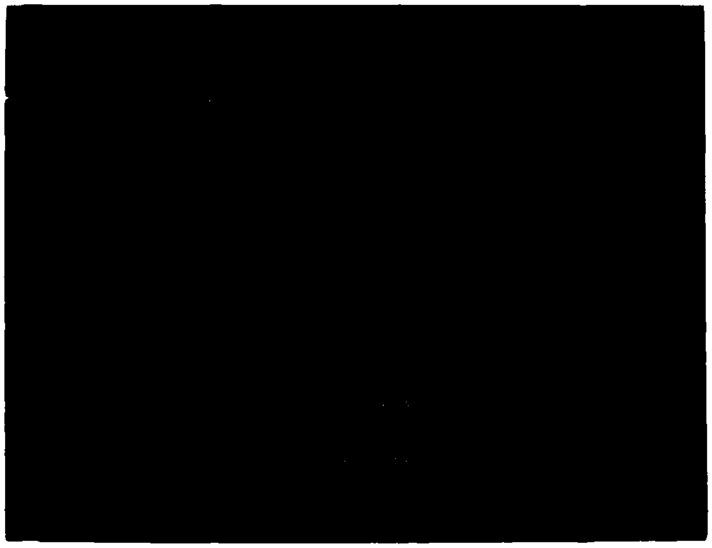


Figure 3-1. An aerial image.

(a) Constraints for a house: compact rectangle

(b) 1. Constraints for a neighboring house

    1.1. Overlaps with the area of interest

    1.2. Satisfies the constraints for a house

  2. Constraints for a neighboring road piece

    2.1. Overlaps with the area of interest

    2.2. Satisfies the constraints for a road piece

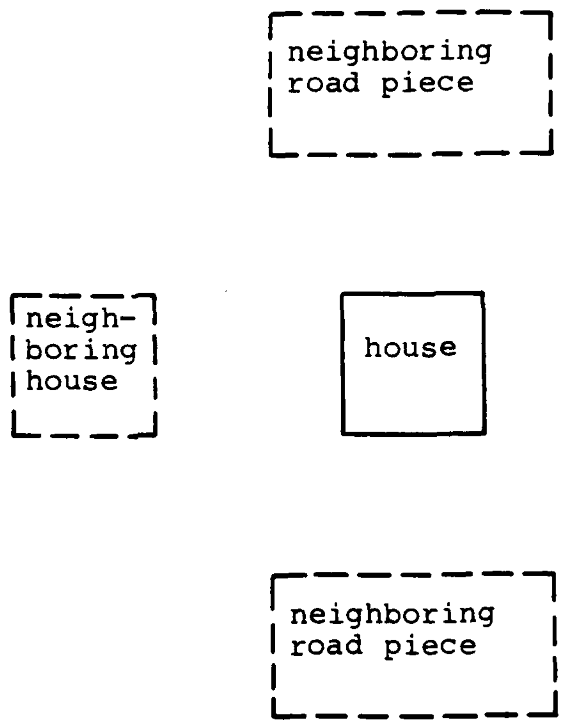


Figure 3-2. Constraints for a house and the spatial relations between a house and other objects.

(a) Constraints for a road piece: elongated rectangle

(b) 1. Constraints for a neighboring road piece

1.1. Overlaps with the area of interest

1.2. Satisfies the constraints for a road piece

1.3. Adjacent to the existing road piece

1.4. Has width compatible with that of the existing road piece

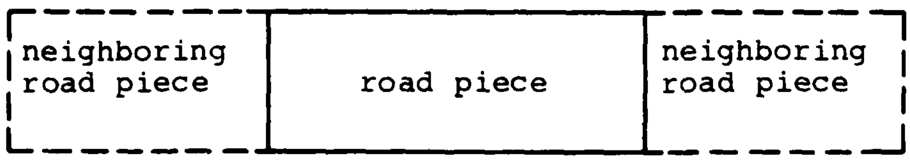


Figure 3-3. Constraints for a road piece and the spatial relations between a road piece and other objects.

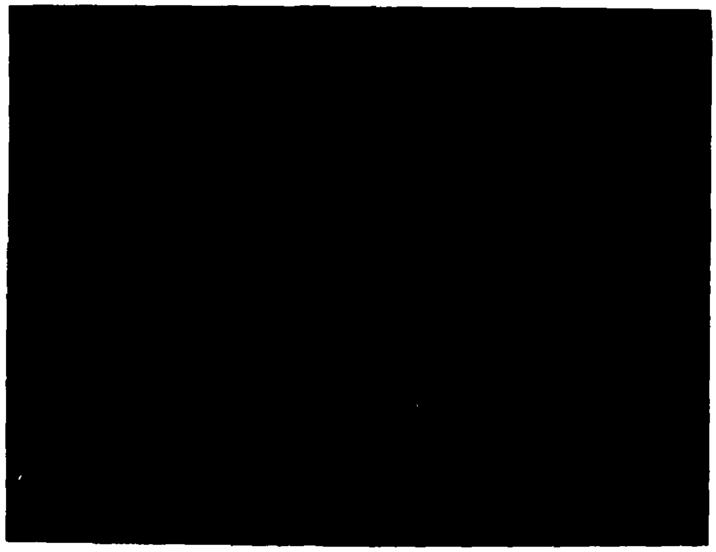


Figure 3-4. Original image (bottom) and initial house instances (top).

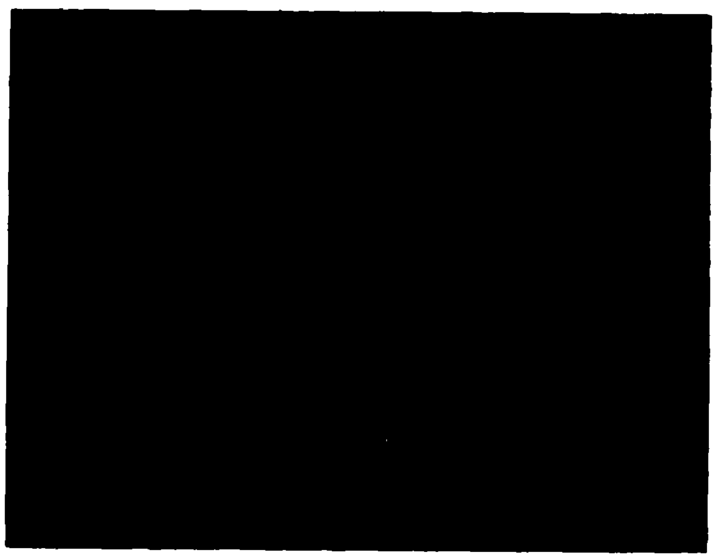


Figure 3-5. Original image (bottom) and initial road piece instances (top).

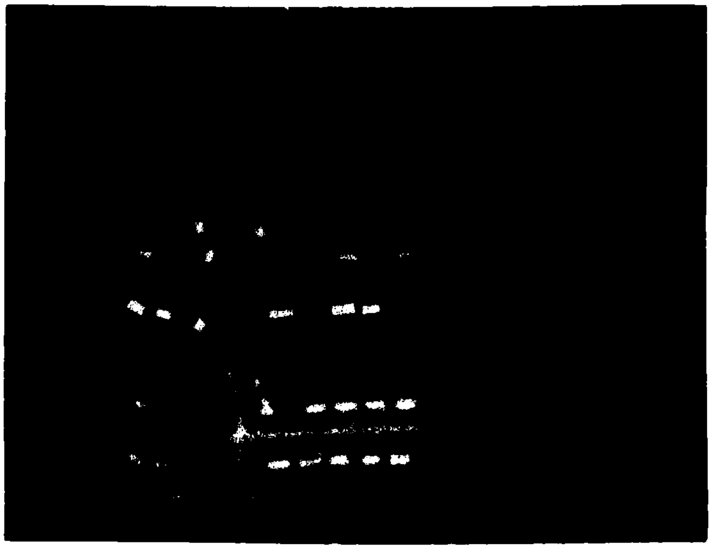


Figure 3-6. Original image (bottom) and all house evidence (top).

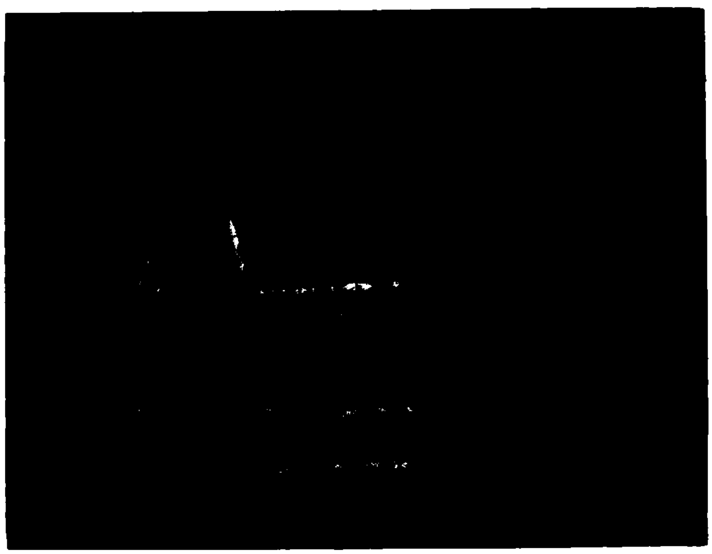
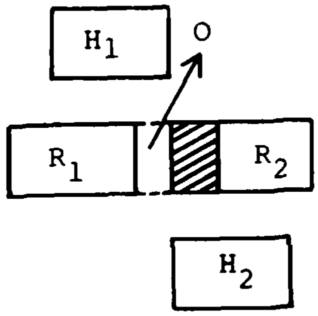


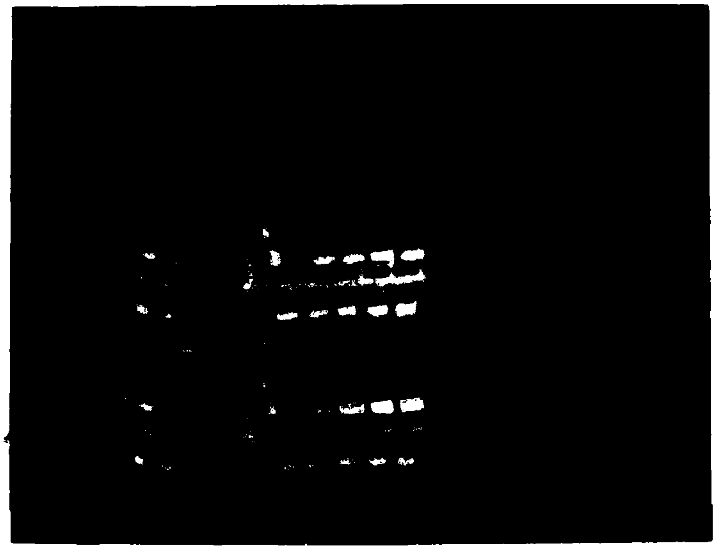
Figure 3-7. Original image (bottom) and all road piece evidence (top).

(a) Selected situation overlayed on the original image (bottom) and the target region of the action overlayed on the original image (top).

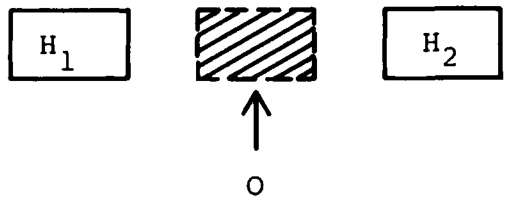


(b) A depiction of the situation.

Figure 3-8. A situation.



(a) Selected situation overlayed on the original image (bottom) and the target region of the action overlayed on the original image (top).



(b) A depiction of the situation.

Figure 3-9. A situation.

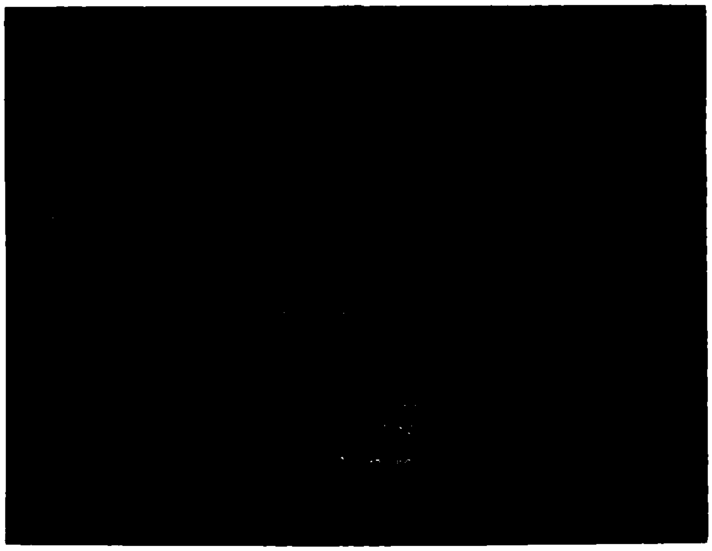


Figure 3-10. Original image (bottom) and all the house instances (top) when the situation is selected.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFOSR-TR-84-0178	2. GOVT ACCESSION NO. A139874	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) EVIDENCE ACCUMULATION FOR SPATIAL REASONING IN AERIAL IMAGE UNDERSTANDING		5. TYPE OF REPORT & PERIOD COVERED Technical 6. PERFORMING ORG. REPORT NUMBER CAR-TR-28; CS-TR-1336
7. AUTHOR(s) Vincent Shang-Shouq Hwang Takashi Matsuyama Azriel Rosenfeld Larry S. Davis		8. CONTRACT OR GRANT NUMBER(s) F-49620-83-C-0082
9. PERFORMING ORGANIZATION NAME AND ADDRESS Center for Automation Research University of Maryland College Park, MD 20742		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2304/A7
11. CONTROLLING OFFICE NAME AND ADDRESS Math & Info. Sciences, AFOSR/NM Bolling AFB Washington, DC 20332		12. REPORT DATE October 1983 13. NUMBER OF PAGES 38
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

We describe a control structure for building an Image Understanding System. This system can deal with objects with diverse appearances when consistent spatial relations exist between objects. By accumulating consistent predictions originated from existing instances, our system can dynamically reason about what to do in order to construct interpretations of the image. In this paper, we have discussed parts of the proposed system - the representation of spatial knowledge, the accumulation of evidence, the focus of attention

ABSTRACT, cont'd.

mechanism, and the integration of constraints for top-down control.

END

FILMED

5-84

DTIC